AI Development Task

Task: Develop Offline AI-Based English-Arabic Bidirectional Text Translator

Objective:

Build an AI-powered system capable of accurately translating text from English to Arabic and vice versa, including very long inputs, using open-source libraries only. The system must operate fully offline once set up.

Scope & Requirements

1. Core Functionality

- Support bidirectional translation:

  - English ? Arabic

  - Arabic ? English

- Handle very long text documents (several thousand words without truncation or performance degradation).

- Ensure high-quality translation with contextual and grammatical correctness.

2. Offline Capability

- Entire pipeline (model inference and processing) must run without internet access.

- Model, tokenizer, and dependencies must be locally available/downloadable during setup.

3. Open Source Stack

- Use only open-source frameworks and models.

- Suggested libraries/models:

  - Helsinki-NLP/opus-mt models

  - MarianMT (via HuggingFace Transformers)

  - Alternatives like argos-translate can also be considered.

- Do not use any proprietary or paid APIs (e.g., Google Translate, Azure, etc.).

4. Performance & Optimization

- Efficient handling of long inputs (e.g., chunking with reassembly while preserving meaning).

- Minimal RAM/CPU usage preferred but accuracy is prioritized.

- CLI tool or Python script is acceptable, with a simple interface (e.g., translate.py input.txt output.txt --lang en2ar).

5. Deliverables

- Full source code (GitHub repo or zip file).

- Clear documentation:

  - Installation instructions

  - Usage instructions

  - Model download/setup steps

  - Offline usage guide

- (Optional) Sample input/output examples for testing.

Timeline

Deadline: 1 working day from task assignment